CLAYTONE SIKASOTE

https://csikasote.github.io

claytonsikasote@gmail.com | +260-979-084696

RESEARCH INTEREST

I'm interested in speech and language processing. In particular, I am enthusiastic about building efficient and robust speech recognition systems in low-resource and multilingual settings. Additionally, I am also interested in understanding the intricacies of human-machine interactions towards building intelligent multimodal dialogue systems. Presently, I'm focussed on building speech and language data resources for Zambian languages.

EDUCATION

MSc., Mathematical Sciences, Africa Institute of Mathematical Sciences, Kigali, Rwanda   Oct 2019 – **Aug 2020**

- Supervisor: Prof. Antonios Anastasopoulos, George Mason University, U.S.A

- Specialization: **Machine Intelligence**

MSc., Information Technology Management, Binary University, Puchong, Malaysia   Apr 2013 – Oct 2014

- Specialization: **Information Technology**

BSc., Computer Science, University of Zambia, Lusaka, Zambia   Apr 2007 – Aug 2011

- Specialization: **Computer Systems Engineering**

TEACHING EXPERIENCE

Lecturer, Department of Computer Science, University of Zambia, Zambia   May 2015 – **Present**

- **Bachelor of Computer Science courses:**

CSC 2011 - Computer Systems	2021/2022
CSC 2702 - Databases & Information System	2021/2022
CSC 3712 - Advanced Databases	2021/2022
CSC 4792 - Data Mining and Warehousing	2021/2022
CSC 4000 - Final Year Project Supervision	2021/2022

- **Postgraduate Diploma in Computer Science courses:**

CSC 5711 - Databases & Information System	2021/2022

Instructor, CISCO Academy, University of Zambia, Zambia   Sep 2016 – Jun 2019

- Instructor of Python Programming,

- IT Essentials and Cisco Certified Network Associate (CCNA) Level 1 and Level 2 CISCO courses.

Teaching Assistant, Department of Computer Science, University of Zambia, Zambia   Feb 2012 – May 2015

- Teaching assistant (Tutor) for Database Systems courses under the Staff Development Fellowship (SDF).

SELECTED PUBLICATIONS

- **C. Sikasote**, K. Siaminwe, S. Mwape, B. Zulu, M. Phiri, M.C. Phiri, D. Zulu, M. Nyirenda, A. Anastasopoulos, *Zambezi Voice: A Multilingual Speech Corpus for Zambian Languages*, INTERSPEEECH, 2023. [*Paper accepted to main conference*]

- **C. Sikasote**, E. Mukonde, M.M Alam, A. Anastasopoulos, *BIG-C: a Multimodal Multi-Purpose Dataset for Bemba*, Association for Computational Linguistics (ACL), 2023. [*Paper accepted to main conference*]

- **C. Sikasote**, A. Anastasopoulos, *BembaSpeech: A Speech Recognition Corpus for the Bemba Language*, Proceedings of the Language and Resource Evaluation Conference (LREC), 2022.

SCHOLARSHIPS AND AWARDS
- **African Master of Machine Intelligence Scholarship**, Kigali, Rwanda — Sep 2019
- **Special Research Fellowship**, University of Zambia, Lusaka, Zambia — Sep 2019
- **Binary University Scholarship**, Binary University, Malaysia — Apr 2013
- **Staff Development Fellowship**, University of Zambia, Lusaka, Zambia — Feb 2012
- **Zambia Higher Education Scholarship**, University of Zambia, Lusaka, Zambia — Apr 2007

GRANTS
- **Co-Principal Investigator**, Multimodal Datasets for the Bemba Language Project — Apr 2021 - Aug 2022

TECHNICAL SKILLS
- **Programming Languages:** Python
- **Database technologies:** MS SQL, MySQL
- **Operating System platforms:** Windows, Linux
- **Machine Learning libraries:** Scikit-learn, Pandas, NumPy, SciPy
- **Data Visualization tools:** Matplotlib, tSNE
- **Deep Learning frameworks:** Pytorch, Tensorflow
- **Speech Recognition Tools:** DeepSpeech, FairSeq

OPEN-SOURCE DATASETS
- **Zambezi Voice** : [Lead researcher and creator] of a multilingual speech corpus containing speech recognition data resources for native African languages of Zambia. [ON-GOING]

  Link: *https://github.com/unza-speech-lab/zambezi-voice*
- **Multimodal Dataset for the Bemba** : [Lead researcher and Co-PI] of a large scale multimodal dataset for Bemba language containing multi-turn image-grounded dialogues amounting to over 180 hours of speech with corresponding Bemba transcriptions and translation in English. The project was funded by *Lacuna Fund*.

  Link: *https://github.com/csikasote/bigc*
- **BembaSpeech** : [Lead researcher and creator] an ASR corpus consisting of 24 hrs of read speech for the Bemba language spoken by over 30% of the population in Zambia.

  Link: *https://github.com/csikasote/BembaSpeech*
- **Multilingual ASR Corpus for African Languages** : [Contributor] to a multilingual speech corpus consisting of over 60 hours of read speech for 19 African languages. I contributed 2 hours of read speech for Bemba language.

  Link: *https://github.com/besacier/AMMIcourse/tree/master/STUDENTS-RETURN*

STUDENT SUPERVISION AND MENTORING
- **Nawiti Simwanza**, Undergraduate, University of Zambia — 2022/2023
- **Felix Kunda**, Undergraduate, University of Zambia — 2022/2023
- **Ngonidzashe Mbita Sikazwe**, Undergraduate, University of Zambia — 2022/2023
- **Stanly Mwape**, Undergraduate, University of Zambia — 2021/2022
- **Kalinda Siaminwe**, Undergraduate, University of Zambia, — 2021/2022
- **Nawiti Simwanza**, Undergraduate, University of Zambia — 2022/2023

COMMUNITY SERVICE
- **Conference Volunteer**

  International Conference on Learning Representations (ICLR) — Mar 2021
  International Conference on Artificial Intelligence and Statistics (AISTATS) — Apr 2021